80

SECRET

SECRET

CENTRAL REFERENCE FACILITIES

STATUS

(1 November, 1949)

and

OBJECTIVES

C E N T R A L I N T E L L I G E N C E A G E N C Y

SECRET

SECRET

RC
OAD/CR
Ref 7, file 57-98
Policy 1949

CENTRAL REFERENCE FACILITIES

SUMMARY

1. Neither production nor collection of intelligence can be effective without a means by which all available intelligence materials bearing upon a given subject or geographic area may rapidly be listed, and rapidly provided for the analyst or collector. If the analyst does not receive such a service he must devote much time and effort, which might otherwise be applied to actual analysis and production, to searching out and bringing together the source materials which he needs. If the collector does not have such a service there is likelihood that a large part of his collection effort may duplicate collection effort already expended by others.

To provide such a central reference service is the over-riding objective of OCD. There is reason now to believe that we are on the track of a far more effective solution to the problem than has ever, anywhere, been achieved before. It is the object of this paper to describe the problem, to state what has been accomplished, and to define as nearly as may be the objective toward which we are working.

2. The best encyclopedia would be of scant value if its articles were scattered through umpteen volumes without any alphabetic arrangement, and without any index or table of contents. Similarly, the CIA Library would be of little use, even though it contained all the intelligence reports available to the nation, if it had no index system by which it could rapidly list and locate all the reports dealing with a particular subject or area.

The volume of documents to be dealt with is immense. Excluding cables, 15,000 intelligence reports were added to the files in October, 1949, bringing the total to about 420,000. All are ticketed and made accessible by means of simple finder files (see page 10), but these do not give any useful indication of the content of the reports.

True index files (see page 18), giving coded information as to the subjects and areas covered by each report, were commenced in April, 1948, when it became apparent that the Agency might reasonably anticipate success in its endeavor to develop an adequate machine system for the handling and reproduction of such files. The Library's indexing system is now keeping abreast of the

daily intake of about 690 documents, and on 1 November it had indexed about 143,000 of the 420,000 reports in the files. The backlog of indexed materials is being worked, within manpower limits, on a selective basis with first priority assigned to CIA-produced reports and to reports from any source which deal with Soviet Russia and its satellites. The diagrams on pages 19 and 20 show how the document and index files are growing and what their current status is.

3. Good as they are, the index files are not enough. They serve admirably the needs of customers who ask, "What documents have we received in the past month on Costaraguan foreign relations?", and these customers are many, but the ORE specialist on Costaragua has no need to ask such a question. He knows what has come in during the past month, because it has been across his desk, and likely as not he has written an estimate on the subject. What he most often wants is a complete list of all available reports which might provide background for a monographic study on a fairly broad subject. The list which he gets from the Library is likely to be so long that he cannot hope to read all the reports, and it gives him no clue as to which of the reports are of greatest value. What he needs, obviously, is a list which gives him titles of all the available reports plus abstracts of the more important ones. Perhaps the abstracts themselves will serve his needs and, if not, they will at least help him to judge which are the documents most likely to suit his purpose.

The Intellofax system provides the mechanical means by which the Library might give the analyst what he needs, for the faxcards provide space for about 150 words in addition to the title and other identifying data. But how is the Library to winnow the wheat from the chaff? How is it to decide which documents are important, and what it is in them that is important? It is obvious that it cannot do so. Only the specialist can decide which are the important documents bearing upon the field of his specialty, and only the specialist can write competent abstracts of such documents.

In January of this year the NEA Division of ORE generously volunteered to give OCD a hand on this problem, and the specialists on its Turkey desk commenced selecting and abstracting those documents bearing on Turkey which they regarded as most important. There was no assurance that the experiment would work.

It might be that the analysts would find themselves taking on a chore which paid scant dividends, and maybe it would be impossible to work out a classification scheme, integral with that
25X6A of the Library, which would enable OCD to cross-

[redacted]

it could be extended in such a way as to satisfy specialists on other areas. OSI commenced operating on the same trial basis in February, and the Greece desk of ORE in March. OCD meanwhile continued to write abstracts of the "table of contents" type which are needed for publications covering a broad variety of subjects and areas. Actual samples of OCD, ORE and OSI abstract cards are shown on pages 23 and 25.

4. The abstract files, whose evolution is described above, are proving exceedingly useful, and it has been found in ORE that the process of writing abstracts has in itself yielded a number of unforeseen but highly valuable by-products:

(a) Time saved in producing weekly and

25X6A

[redacted]

abstracts.

(b) Training of junior professionals is improved, speeded up, and locked into the production system so thoroughly that it is no longer in danger of being overlooked or indefinitely postponed. The desk chief assigns particular incoming reports to particular juniors for abstracting, they must dig deeply into the reports to write competent abstracts, and each abstract is edited by the desk chief. In this way the output of each junior comes under constant inspection, weaknesses can be pointed out and corrected, and

a high premium is placed on first-rate work,

(c) Technically difficult reports no longer go into the "Hold" basket or files to be attended to when there is more time. They are assigned to a junior, and he is obliged forthwith to scratch around and acquire enough background, technical and otherwise, to deal with the complexities in the report and thoroughly understand them.

(d) File space is saved. The [redacted] desk 25X6A estimates that over 50% of its filing space will be freed, as it is operating on the principle that each abstract should be so complete in itself as to make it unnecessary, in most cases, to refer again to the original report. Hence the reports themselves may be destroyed or sent to the Library.

5. Rapid growth of the abstract files is cause for administrative concern. On November 1st they held about 17,500 abstracts, and they are growing at a staggering rate - about 1,500 in August, 2,000 in September, and 3,000 in October. OCD's typing and classification people are now working overtime to keep abreast, and more hands will be needed if the system is to be further extended. Economies gained by use of the system seem likely to justify extension - as typists cost the Agency less than professionals - but progress will have to be made on a step-by-step basis in any event. Each extension of the system to a new group of specialists involves difficult adjustments, revisions, and expansions in the Classification Scheme, and these can be accomplished only by day-to-day hard work. Any attempt to extend the system simultaneously to all ORE specialists would create chaotic confusion in the Classification Scheme and thus would destroy all chance of developing a system satisfactory to all concerned. Evolution, not revolution, is called for.

6. OBJECTIVES. The extent to which the index and abstract files are now being used by all offices of CIA, and by the IAC agencies, seems clear evidence that we are on the right track. Early in 1950 we hope to have two Faxcard Transmitters in OCD, and five Receivers located at strategic points in M and Q buildings. All analysts in the two buildings will then have, within easy reach, equipment somewhat similar to teletype which can produce for them upon request lists of all documents in the

Library which bear upon a given problem, together with abstracts of those which have been selected as most important by analysts of their own or other CIA components. The index files give us good substantive control of the current flow of intelligence documents from all sources, and the abstract files are beginning to give us even better substantive control of all important intelligence materials - whether from the usual intelligence sources, or from the foreign press, or from any other source known to the analysts concerned.

Because of the selectivity being exercised in building up the abstract files, we are forging a tool which in years ahead will enable us to drain off from the Library those documents which are of scant importance, thus making room for then current inflow and ensuring that reports of real importance are held available.

It is quite possible that the central reference system being built by CIA will ultimately prove the most important central intelligence service which the Agency provides.

G L O S S A R Y

7. The following definitions, some of them novel, must be understood for an intelligent reading of this report:

Abstract	A brief summary, extract from, or description of an intelligence document. Each OCD abstract is designed to give a broad description of the different kinds of subject matter to be found in a given report, and such abstracts are prepared only for documents covering a wide variety of subject matter. Each ORE or OSI abstract synthesizes the content of a report which the analyst regards as important from the viewpoint of his own specialty.
Abstract Faxcard	A faxcard (see definition below) which shows in print not only the essential index

information for a given document but also an abstract. Such cards are prepared only for documents which are regarded as having considerable importance, and some give abstracts of cables, foreign newspaper articles, or other materials which are not indexed by OCD but which are selected as important by ORE or OSI analysts. These cards are coded and punched for cross-reference purposes as are the Index Faxcards. For samples see page 23.

Abstract Files

The complete file of Abstract Faxcards, including on 1 November 1949, a total of 17,512 abstracts. Of these, 6,095 were written by OCD, 2,564 by the Turkey desk of ORE, 2,709 by the Greece desk, and 6,144 by OSI. A graph, showing the growth of the files, will be found on page 26.

Classification Scheme

This has nothing to do with security classification: it is the 6-digit numerical framework developed by OCD with IAC, ORE and OSI cooperation which provides the code-numbers used in cross-referencing the subject matter of intelligence reports. The term is in universal use in this context and it is unfortunnate, for those in the "sensitive" agencies, that they are obliged almost invariably to use "classify" in a limited and bastard sense. Instead of "Classification Scheme" we might use "Categorization Scheme", but this is ugly and unwieldy.

The CIA Classification Scheme is under continual revision to meet the needs of CIA analysts. For a description of its evolution, and the problems involved, see page 13 et seq.

Coding

The process of determining under what subject headings an intelligence document should be cross-referenced, and of indicating the appropriate numbers from the Classification Scheme. (NB: The Civil Service Commission

distinguishes between "classifying", which is the first and professional half of the operation, and "coding" which is the second half and can be done by anyone. The distinction has no usefulness for any jobs performed in CIA, though it is valid in such offices as the Bureau of the Census which must deal with an immense volume of data all of which must be handled in precisely the same way.)

Faxcard — The faxcard, or Intellofax Card, is an IBM punchcard of standard shape and dimensions which bears on its face up to 200 words of printed information (including title and abstract), and which at one end bears the coded and punched classification numbers for cross-reference by machine techniques. These cards are sorted, selected, and arranged by standard IBM machines; and the printed information on the selected cards is thereafter transmitted and reproduced by Intellofax. Sample faxcards from the Index File are shown on page 18, and from the Abstract File on pages 23 and 25.

Faxcard Transmitter — A machine, at least two of which will soon be located in OCD, which can automatically take a file of selected faxcards and transmit electronically the printed information which they bear on their faces. Transmission is by cable to a Faxcard Receiver.

Faxcard Receiver — A machine, at least five of which will soon be located at strategic points in M and Q buildings, which can receive printed information from the Faxcard Transmitter and reproduce this information on a continuously flowing tape somewhat similar to, though smaller than, the wide tape in a teletype receiver.

Index Faxcard	A faxcard which shows in print only the title, originator, classification, date, and number of pages in an intelligence report. There may be one or many for a given report: when there are many, each is coded and punched differently for cross-reference purposes. These cards are now being prepared by OCD for all incoming intelligence documents except cables, and it is intended that they shall serve as a complete and fully cross-referenced index to all intelligence reports received or issued by CIA. For samples see page 18.
Index Files	The complete files of Index Faxcards prepared by OCD. Now being kept up on a daily basis, and up-to-date on current inflow. Backlog of unindexed reports is being indexed as opportunity and manpower permit, with first priority being given to reports from any source which deal with Soviet Russia and its satellites.
Intellofax	The system, combining IBM machine techniques and facsimile reproduction techniques, which has been developed by Finch Telecommunications, Inc. and OCD for the use of CIA. The two essential items of equipment are the Faxcard Transmitter and the Faxcard Receiver.
Intellofax Files	All files prepared on faxcards. The term embraces both Index and Abstract files.

MAGNITUDE OF THE PROBLEM

8. The flow of paper into CIA is tremendous, and it takes considerable manpower to keep the current moving. Fortunately, only a portion of the flow needs to be considered for reference purposes, as only a small number of documents is received in single copies.

The figures given in this paper refer only to intelligence documents and to separate titles. They do not include cables, for reasons discussed

below. Finally, the figures must be regarded as estimates only and subject to a varying margin of error. Manpower to handle the incoming documents has never been great enough to permit exact daily counts of the reports coming in from many different sources and such counts, in any case, would still be inadequate for two reasons: (1) many documents filter down to the Library long after they have come to one or another CIA office through direct and informal channels, and they are often indistinguishable from other documents of similar character which have already been counted, and (2) it is often difficult to decide just how a particular item should be counted. An MA report, for example, which is in fact no more than a letter of transmittal with two important enclosures might be counted as one, as two, or as three intelligence documents. No doubt it would be possible, with sufficient manpower, to set up a system which would provide exact counts; but the cost of such an operation would be greater than any benefits derived from it.

9. Cables, telegrams, and the like are not considered a part of the central reference problem, and they are not included in any of the estimates used in this paper. The reason for this is that cables and telegrams, although they are vital for operational and current intelligence purposes, seldom have any value a month after their initial receipt. Those of scant value even for current intelligence estimates may be forgotten, and those of considerable importance are likely to be followed up by fuller and more detailed reports. Hence it was early decided that no attempt would be made to index incoming cables, and that no records would be kept on them save those needed for security controls. It has been found that the Library almost never receives a request for a cable more than six months old, and therefore cables of greater age than this are destroyed to make room for current inflow. If ever it becomes important to get a copy of an old cable it will be possible to do so through the responsible agency, as each such agency is obliged to keep a complete file of its own cable traffic for administrative purposes.

Although, for the reasons stated above, it would be uneconomical to index the cable flow, it has been found in practice that ORE analysts find it desirable to abstract a good

many of them. "When the news is sufficiently important to send by cable, it is likely to be sufficiently important to abstract." This is the view expressed by ORE specialists, and it should be noted that the abstracting scheme provides the means by which important cable materials go into the reference files while unimportant cables are discarded.

10. The volume of intelligence reports in Library files on 1 November, 1949, was about 420,000.

The inflow began in 1946, soon after CIG was organized and had found working quarters, and it increased erratically as first one and then another agency began to contribute its reports. Though no exact count is possible, it is estimated that about 30,000 of the Library's documents came in during 1946. In the ensuing years volume stepped up gradually, and is still increasing. Average monthly intake in 1947 was about 9,400; in 1948 about 11,750, and in the first ten months of 1949 about 13,000. These figures suggest that monthly inflow may finally level off at about 16,000 or 17,000, but there can be no assurance that this will in fact occur. Reduction in field collection effort by the Defense agencies may cause a decline in volume, while increased reporting by State and such agencies as ECA might cause a sharp increase.

It is known that a substantial number of the 420,000 documents, perhaps as many as 10%, need never be indexed for reference purposes. These are, for the most part, letters of transmittal covering enclosures which may or may not have been received by the Agency. Such documents must be filed with the others, even though they have no intelligence content, in order that they may be located upon request before we embarrass ourselves, for example, by asking Army to send us report number so-and-so and being told in return that we have already received it.

11. The nature of the reference problem is best described by categorizing the questions which the Library must answer.

(a) Simplest is the request for a specific document. An analyst asks for Report No. 555/49 from the Military Attache in Belgrade, or perhaps for a report which has the CIA number 123456.

Simple finder files, of the type kept by every library, tell the librarians whether or not the report has ever been received. These files are on 5' x 3'

slips of paper, one file arranged by name of the originating office and its serial number, the other arranged by CIA number. If it turns out that the report has come in but cannot be found in the MA Belgrade file, then the charge-out file kept by the Circulation Desk will tell who has borrowed it from the Library. If there is no record of the report's having come in to CIA, then the Library arranges to borrow a copy from the Army's library. If the latter proves recalcitrant, then the problem is turned over to a Liaison Officer and he commences negotiations in the Pentagon.

(b) The reference problem really begins to get tough when someone asks for a list of all reports which have come in during the past month, and from any source, which deal with the foreign relations of Costaragua. Many requests of this nature are received every day, and they generally come from agencies outside of CIA, or from OO, OSO or OPC. They seldom come from ORE or OSI, because each specialist in those offices has presumably had on his desk all the reports received during the last month which have a bearing on his specialty.

The Library is now able to give fast and reasonably good service on requests of this nature by means of its Index Files, described on page 18.

(c) The third type of question is that which comes in most often from an ORE or OSI specialist, or sometimes an analyst in one of the other offices or agencies, who has been directed to prepare a full length study on a specialized subject. What he asks for, perhaps, is a list of all available documents which touch upon relations between politicians and the military in Sylvania. He would like, if possible, to have brief descriptions or abstracts of the more important documents on the list in order that he may decide, without much time-consuming research, which are the ones most likely to be of value for his particular purpose.

In academic life, where time is of no great consequence, the researcher must perform this sort of service for himself. He gets from the library's card catalogue a list of the dozen or more volumes which deal in general with his problem, then he consults the bibliographies given in each of them for additional titles. Gradually he

builds up an imposing bibliography for his own study, and over a period of months he sifts the books one by one for the chapters, sections, or statistical tables which have a specific bearing on his job.

National intelligence studies cannot be handled in this admirable but time-consuming and expensive way. The researcher is told: "Get the right dope, and get it fast!" Deadlines must be met, and CIA must face the problem squarely. If it can find a solution it will be performing a service whose value to all the agencies can scarcely be overestimated. The Abstract Files, described on page 21, now seem to be our most promising approach.

12. The volume of reference requests which the Library must answer is considerable, and it is increased by each new accretion to CIA's research or operating staff. Up to November 1st the Library had answered some 685 requests with the Intellofax system, and had produced an average of 130 references for each request.

How long would it take, by usual Library methods, to search out 130 documents bearing upon a particular subject and to get the list of titles typed and into the researcher's hands?

Yesterday the Library received 10 different requests for lists of documents bearing upon 7 different countries and 10 wholly different subjects - ranging from inorganic acid production in one country to the relations between another country's military and its politicians. The requests came from ORE, OSI, OO, OSO, the Army, and the National War College. All 10 were answered before closing time, with lists which ranged from 4 titles in one case to 178 in another. It would have taken weeks of file-searching and typing to have answered them in any way other than Intellofax.

13. Our attack on the problem must be two-pronged: mechanical, and substantive.

Thousands of documents cannot be combed and sifted rapidly even by a large staff of librarians and clerks. A machine system must be found to do the job, and it must be able to print the titles and abstracts which it culls from the files. The Intellofax System, described on page 15, will do this.

The machine system is worthless unless the code numbers by which it operates are derived from a classification scheme which is adequate, from the user's standpoint, to break down and

categorize the subject matter in our thousands of documents. If we can't assign a code number to what we want, then we can't call on the machines to deliver.

CLASSIFICATION SCHEME

14. To develop a classification scheme which will meet the needs of all concerned is just as important, and even more difficult, than it has been to work out the mechanics of our system. Tough semantic questions enter into this problem, and they are inflamed by the belief of every specialist that he and he alone is competent to classify the field of his own specialty. It is noteworthy that no two such specialists, even when working in precisely the same field, can come to 100% agreement on how their data should be categorized.

The nature of the problem is best shown by example. Here are two different ways in which we might code aluminium cooking pots:

100. NON-FERROUS METALS
120. ALUMINIUM
123. End Products
123.4 Consumer Goods
123.45 Household Utensils
123.456 Cooking Pots

900. ECONOMY
980. INDUSTRIAL PRODUCTION
987. Consumer Goods
987.6 Household Utensils
987.65 Cooking Pots
987.654 Aluminium

15. The importance of this problem was immediately recognized by the CIA Library which, in 1947, commenced study of the many existing schemes of classification used by the IAC and other government agencies. It was clear at once that the Army's BID system was useless for coding naval subjects, and that the Navy's Monograph Index Guide would not serve for subject matter of primary interest to the Army. The same applied to all the other schemes, including those of the State Department, the Air Force, the Library of Congress, and the Research and Development Board. Round table conferences were then held with representatives of these agencies, in the hope that they might be brought to agreement as to what system CIA should use. This proved useless, as each

insisted that CIA should adopt his system lock, stock and barrel. Thought was then given to the possibility of building up a composite scheme, which would base its military subdivisions on the Army's BID, its naval categories on the Navy's MIG, and so forth; and unilateral discussions were held with representatives of the several agencies in the hope that each might agree as to how those sections of CIA's scheme which were of prime interest to his agency should be handled. This also proved futile, as each representative took a cosmic view of the fields which were of primary interest to his agency, and argued that the whole structure of intelligence would be imperilled by any deviations from his scheme.

16. Time was marching on, but the classification scheme was not. We had gotten nowhere, and it was painfully evident that we never would get anywhere by interdepartmental consultation. Even a bad start would be better than none, so the Library set about drawing up its own scheme with the help of a classification specialist from RDB who had a very keen interest in the problem, who had been an intelligence officer himself, and who had shown readiness to work on the CIA scheme in an objective and disinterested way. The framework decided upon was eight-fold:

100 - World Politics
200 - Army
300 - Navy
400 - Air
500 - Weapons and Scientific Warfare
600 - Science and Technology
700 - Economics
800 - Cultural and Social Forces

Each of the eight categories was broken down to provide, as nearly as possible, for the needs of the agency chiefly concerned - the Army, Navy and Air sections following closely the patterns developed by the three services for their own use. The separate sections, except Army, Navy and Air, were then worked over in detail with the ORE units chiefly concerned (including the then Scientific Branch) and revised. Next the entire scheme was published in order that it might be tried out in practice, and the Library commenced applying it to its intelligence documents. Many further revisions were found necessary, and they were made on a day-to-day basis as the need for them became apparent. Happily, the scheme as initially drawn up proved sufficiently elastic to encompass many unforeseen requirements.

17. Present status. The classification scheme in its present form has been applied to the 143,000 intelligence documents which have been indexed for the Intellofax System, but the process of revision and expansion is still going on, and it will likely continue for a long time. The abstracting program (see page 21) has provided an opportunity for the Library to get pertinent sections of the scheme scrutinized on a daily basis by ORE analysts of the Turkey and Greece desks. In consequence there is now assurance that the scheme does meet the needs of those specialists, and since it has been made to meet their needs there is reasonable assurance that it can be made to meet the needs of specialists on other areas.

18. Objectives must be realistic. It must be recognized that no single classification scheme will ever be 100% acceptable to all analysts. Also, since the Library must provide service to all offices of CIA and to many in the IAC agencies, it is evident that we cannot afford to develop a variety of different classification schemes - each designed to please a particular unit in ORE or OSI. When we call upon the index files to produce the titles of all reports bearing upon Costaraguan rubber we must be able to do it by searching a single segment of the files - not a dozen segments, each arranged according to the views of a different office or analyst. We will do better to develop one scheme which is 90% satisfactory to all customers than to develop 99 schemes each of which is 99% satisfactory to a single shop. Since we cannot hope that we will ever do better than 90% in satisfying all who are concerned, we must anticipate that we will always have complaints to deal with.

INTELLOFAX SYSTEM

19. Usual library procedures, designed to locate the books dealing with a particular subject, would not serve for CIA. The usual library system is to type up subject cards of the same size and shape as the author and title cards. Thus a book on music is represented in the catalog by three cards filed alphabetically - one headed perhaps "Koussevitzky, Sergei", another "The Bull Fiddle", and another "MUSIC - Instruments - Bass Viol". This enables a visitor or a librarian to find out what books the library holds on Music, even though

he doesn't know the name or title of any such books in advance.

The trouble with this system is that it requires the typing and hand-filing of a vast number of cards - if the breakdown by subject and area is to be sufficiently detailed to meet a researcher's needs. The Library of Congress operates as fine a file as could be designed along these lines and, in consequence, researchers from all over the world may be seen there every day searching through the cards and jotting down the titles, authors, and catalog numbers of books which they need for their work. In CIA we are serving a large corps of specialists, each of whom has highly specialized needs, and we must index about 690 documents each working day. It is conceivable that an army of classifiers, typists and file clerks could handle the job in this generally accepted manner, but the cost in space and manhours would be tremendous. Also, even when the job was done, it would still be necessary to search the files by hand and to type the lists of titles and document numbers.

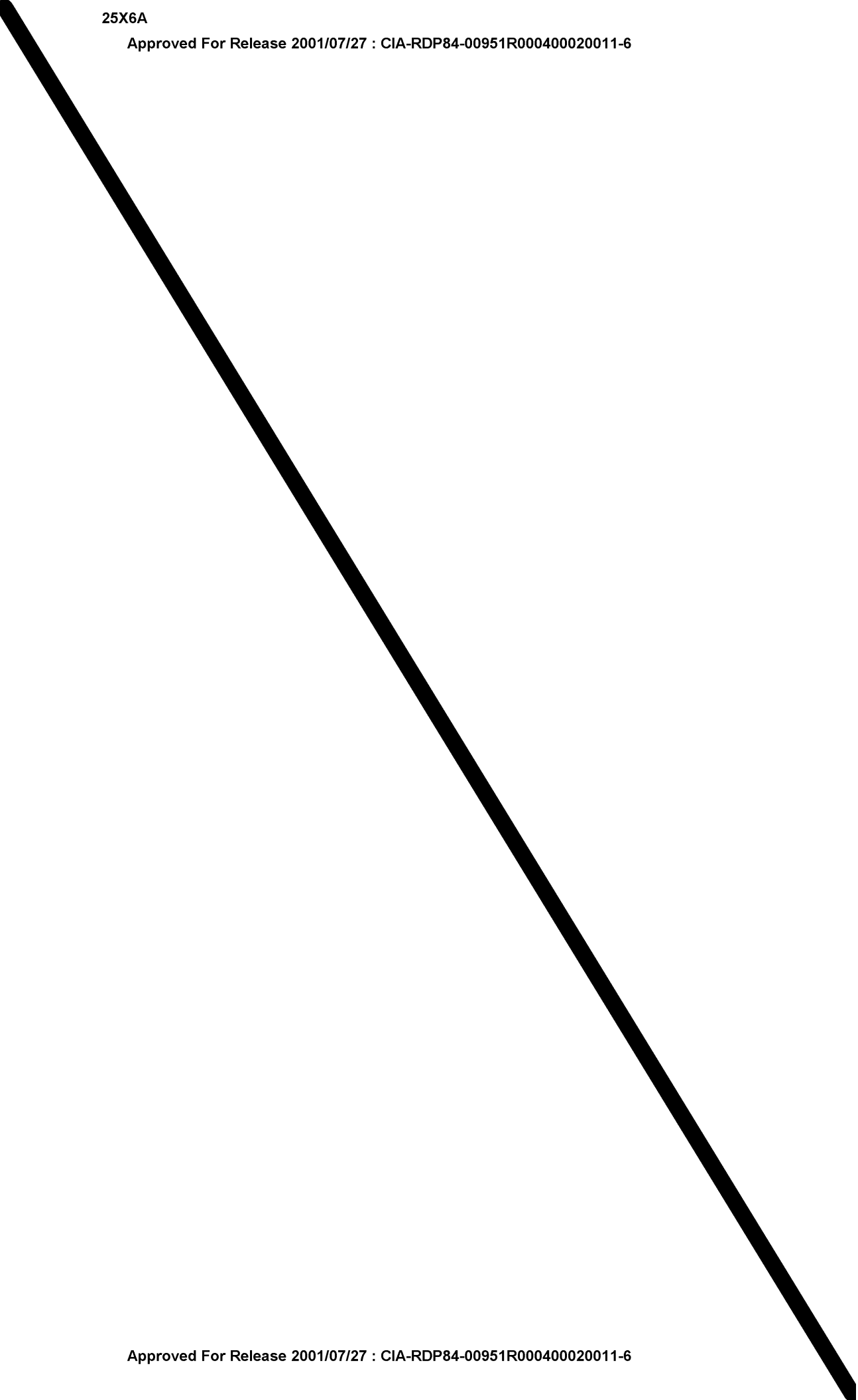
20. Machine techniques, very evidently, offer the only hope of a satisfactory solution - one which will provide both rapid selection of the needed cards and rapid reproduction of the information on them - but when CIA was first organized there did not exist any machine system which was adequate. Standard IBM cards, although they can be selected and arranged in any desired order with all the requisite rapidity, cannot be made to print more than 43 letters and 37 digits from each card. Obviously this is not enough for author's name, title, and other necessary data. Hence, when Wright Field tried to use standard IBM techniques for this purpose, the experiment proved to be so costly that it was never tried again. It was found that to print authors' names and titles, plus a small amount of abstract information, from four to a hundred cards were needed for each entry in the files. The cost in manhours, space, and machine rentals proved to be astronomical.

21. The Intellofax system, now being developed by Finch Telecommunications and OCD's machine people, enables us to foresee in the near future a dependable machine system which will really handle our job. The subject cards in our index and abstract files will be standard IBM punch-cards, punched for as many cross-reference purposes as may be needed, and hence available for rapid selection and arrangement by standard IBM machine equipment; and they will bear on their faces as many as 150 or 200 printed words. The

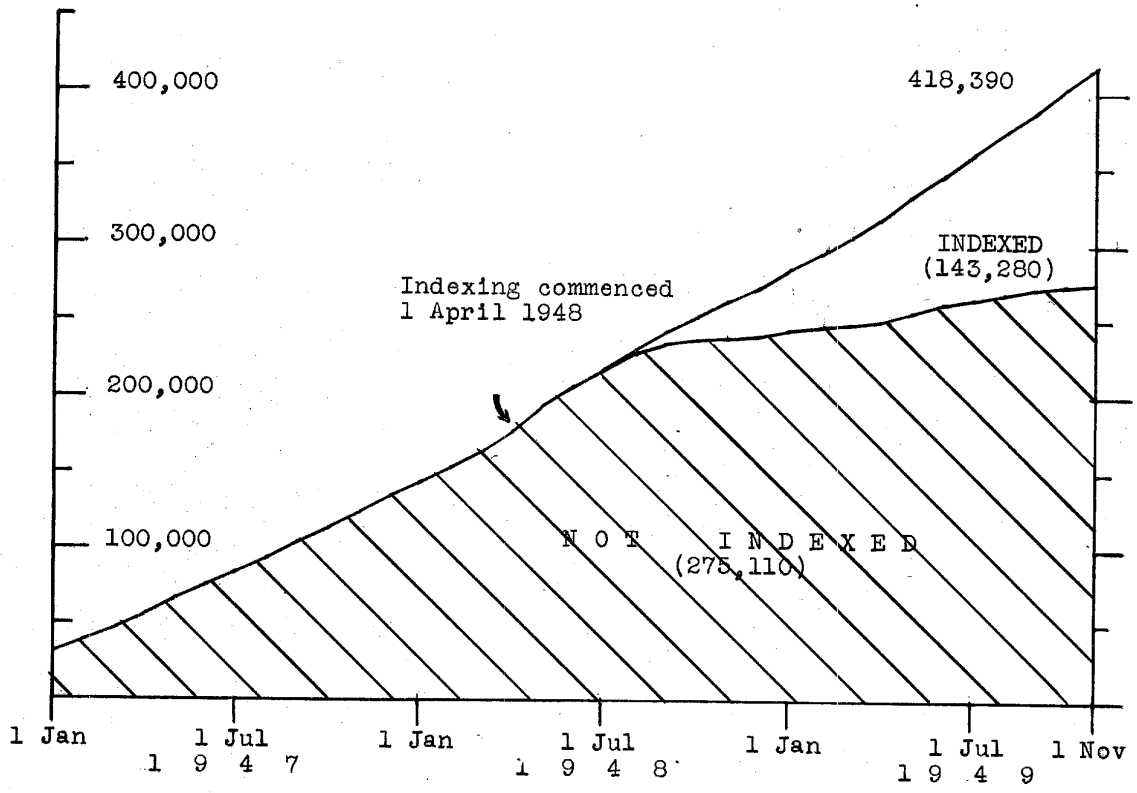
Intellofax equipment will take any number of these cards from the IBM sorter or collator, and will reproduce the typed words on their faces. Thus the analyst will not need to search through the cards in our file: the machines will do that for him, and far faster than he could do it. And no typist will have to copy down the data shown on the selected cards, for the Intellofax will do this automatically.

22. Present status of the Intellofax is as follows: Pilot models of the Faxcard Transmitter and Faxcard Receiver have been completed, and are sporadically in operation. They are far from perfect, and are by no means 100% dependable, but they do produce legible results. We can be sure that we are on the right track, and that before long we will have dependable equipment. Additional transmitters and receivers, with the changes and improvements so far found necessary, are under construction. We hope early in 1950 to have two transmitters located in or near the Library, and five receivers located at strategic points in M and Q buildings where they will be most accessible to analysts. The chore of getting our documents coded and indexed on faxcards is being handled by the Library, and is going on apace. Interim service, pending installation of dependable transmitters and receivers, is being given to all customers by allowing them to scan the actual faxcards selected by IBM sorter from the files in response to their requests.

23. Objectives. Our first objective is to get the Intellofax equipment perfected and installed in M and Q buildings. The chief worry is dependability for, like all newly invented machines, the transmitters and receivers are sure to develop unforeseen crankiness when placed in continuous service. When dependability has been achieved we will wish (a) to develop security control of transmissions, perhaps by use of scramblers, in order that receivers may be placed in buildings apart from M and Q, and (b) to seek improvement in legibility of the characters reproduced on the receiver tape. Both developments will be difficult, but they are not technically impossible.

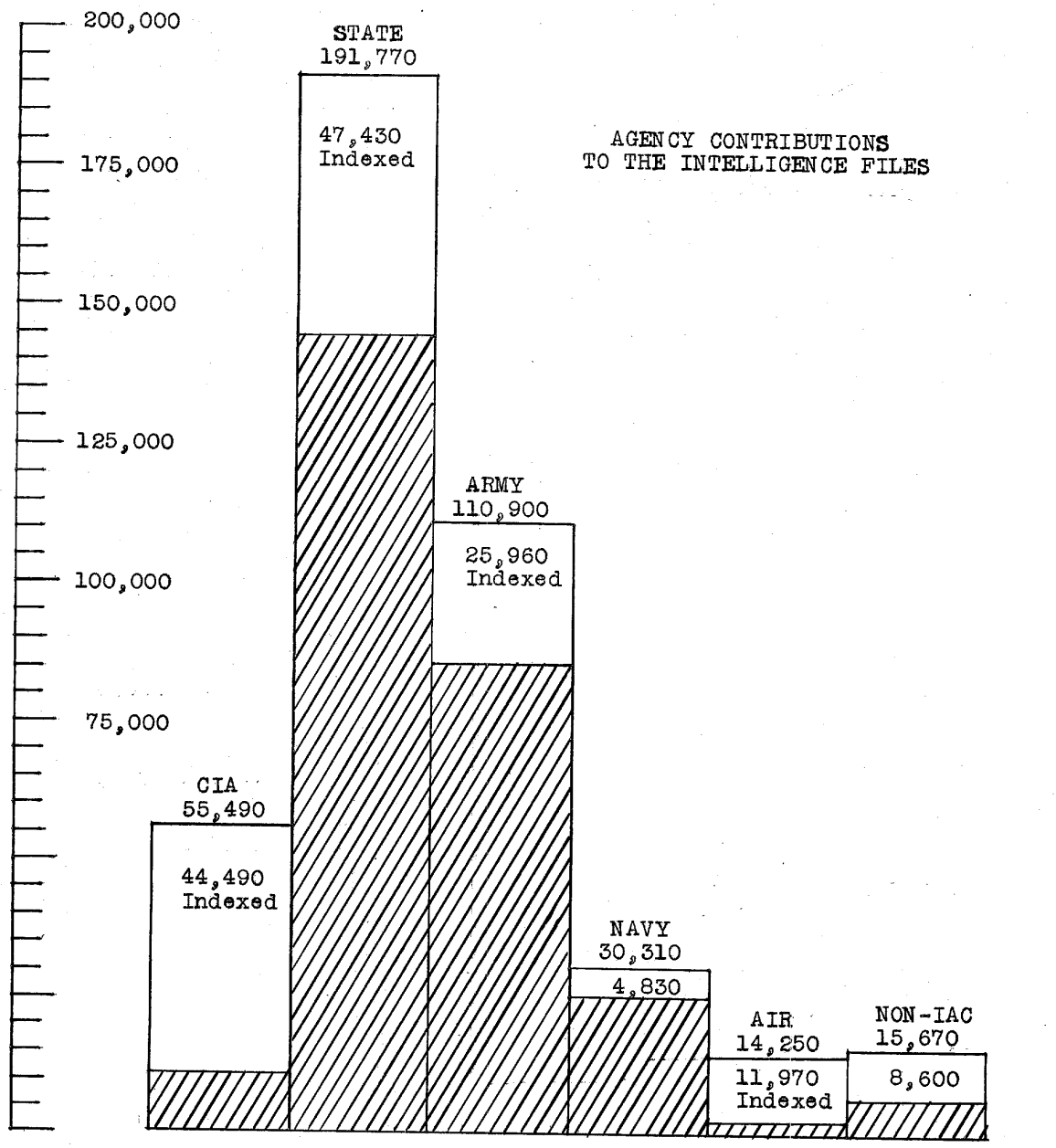


INTELLIGENCE DOCUMENTS FILES
Cumulative Growth to 1 November 1949



problem. The start was slow, as unforeseen difficulties arose in the Classification Scheme as soon as it was put into use, and revisions were frequent. It was not until October that volume was stepped up to a point at which it was keeping pace with the flow of incoming documents.

It will be noted that the cumulative growth curve of the unindexed documents shows a continued uptrend in the graph above. This is accounted for by the number of incoming documents which do not need to be indexed. We do not have sufficient statistical data at present to make a useful estimate of their quantity, but we do know that it is considerable. In spite of the inference which might be drawn from the above curve the Library is, today, keeping pace with inflow; and, as shown on the following page, it has made some inroads on the backlog of CIA-produced documents, and of documents from other sources which deal with Soviet Russia and its satellites.



25. Present status of the Index Files is shown by the above diagram, and it will be noted that most of the documents produced by CIA have now been indexed. The relatively high proportion of indexed documents from Air Force and Non-IAC

agencies is accounted for by the fact that contributions received from these sources have been relatively heaviest in recent months, and they have thus been indexed as part of inflow.

26. Objectives. The Index Files provide the answers to such questions as:

"What documents received during the past month refer to such-and-such a subject and area?"

The question defines the major objectives of the Index Files: (a) that they be kept up-to-date, and (b) that they encompass all intelligence documents coming in to CIA or produced by CIA. By and large, both objectives are now being achieved - for the Library is indexing all incoming intelligence reports within 24 or 48 hours of their receipt. One copy of each report received in multiple copies goes direct to the Library for indexing and file, while the other copies go simultaneously to the analysts concerned. Single-copy documents get priority treatment, in order that they may be duly indexed and transmitted to analysts without delay.

Candor requires an admission that the process is not yet on a wholly solid footing, and that it sometimes bogs down. Delays in handling the incoming documents may be caused by sudden and unexpected increases in the day's take, by a rash of sickness or holiday leaves in the staff, or by the occasional necessity of diverting some of the people to Library jobs of even greater urgency which, for one reason or another, show signs of breakdown. Chain reactions are not uncommon in a unit such as the Library which must keep a heavy volume of paper moving, and which has only the bare minimum of personnel to do the chore. There is bound to be trouble when the people in Room 1048, Tempo M, receive upwards of 20 mailbags in one working day instead of the usual 13.

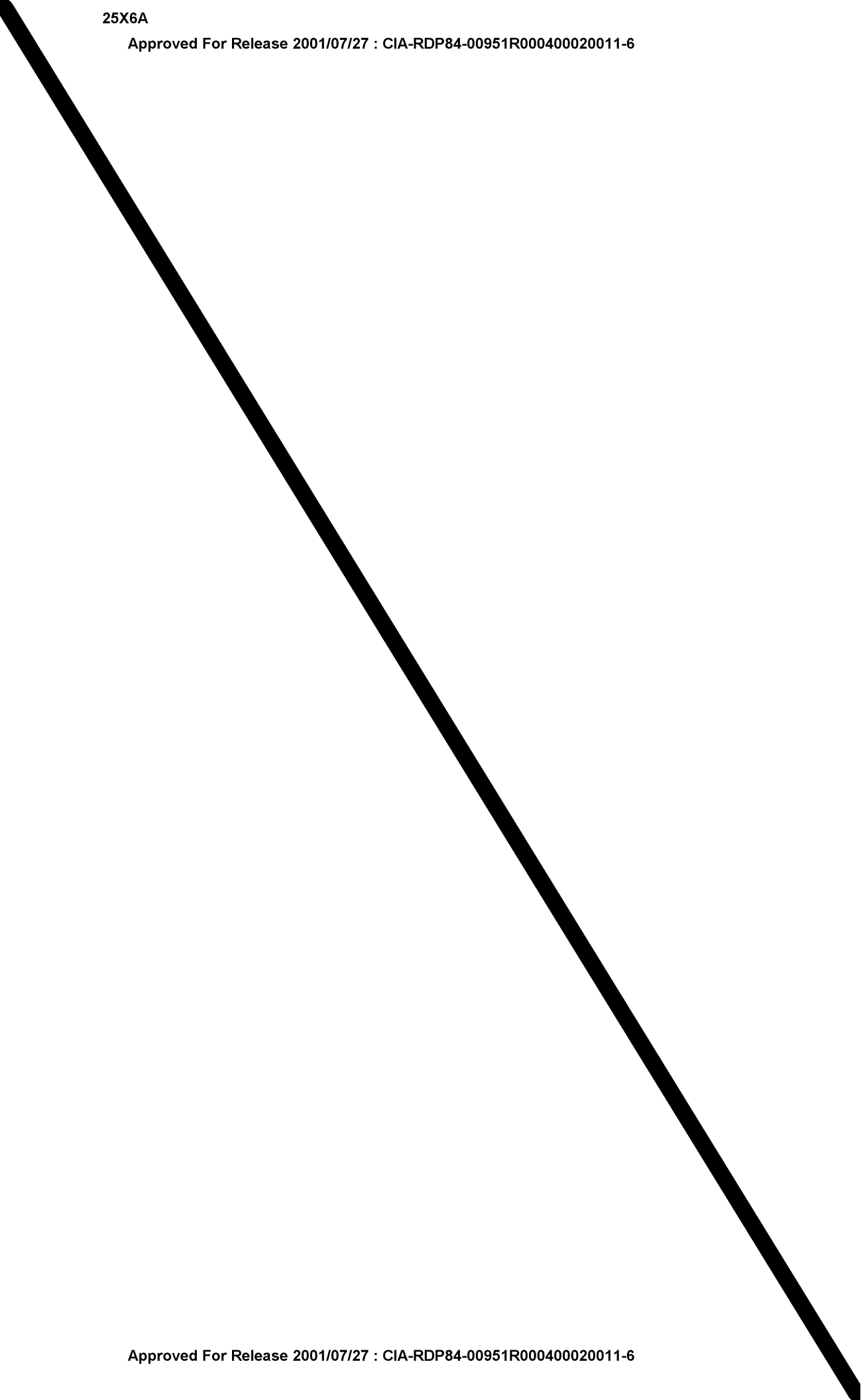
ABSTRACT FILES

27. The Index Files provide an exceedingly useful service for intelligence people, as is adequately demonstrated by the extensive use now being made of them by all offices of CIA and by the IAC agencies. But they have one

serious weakness - they do not differentiate between documents of very slight value and reports of high importance. The analyst who receives a listing from the Index Files must use his intuition in judging which titles look likeliest for his purpose, or else he must ask for all of them and look them over. It would obviously be a great help to him if the important titles in his list were followed by competently written brief abstracts or descriptions of content. He could then make a more rapid and more accurate selection of the documents likely to be useful and, there being fewer of them, the Library could get them from its files and make delivery in much shorter space of time. In a good many instances, perhaps, the abstracts themselves would fully meet the analyst's needs.

But who is going to decide which documents are important, and who is going to write the abstracts? Librarians, clearly, should not evaluate or abstract a report on Panamanian politics or one on malaria controls in Indo China: such work must be done by specialists in the subject matter. But it is equally clear that these same specialists should not be called upon to write descriptions, or abstracts of the table-of-contents type, for periodicals containing articles on many different subjects or areas, for statistical handbooks, or for such other publications as contain a diversity of material but have titles which in themselves are meaningless.

28. In April of 1948, when the indexing job was first started, the Library simultaneously commenced preparing abstracts of the table-of-contents type and also descriptive abstracts of documents which were borrowed from other agencies but which could not, for one reason or another, be reproduced and held permanently in CIA. It could be assumed that these loan documents were important, because they were borrowed in response to specific requests from ORE researchers, and it therefore followed that they should be fully described in the Intellofax files. Thus a document which one researcher had learned about, perhaps by chance, and which had likely been hard to find and borrow, would be made known to all researchers working on similar problems and would be easy to find if needed again. By the end of 1948 some 3,800 abstracts had been prepared by the Library and entered in the files, but no



tested in practice and the Library had no assurance that it would be genuinely satisfactory for research analysts. It was certain (a) that the scheme had been painstakingly drawn up, and that criticism and help from the specialists had been lavishly supplied and gratefully incorporated, and also, (b), that the scheme was in use and was receiving quite as much favorable as unfavorable comment from CIA and IAC customers. But hit or miss comment was not enough, and the Library felt urgent need to work closely with one or another group of research analysts in order to spot, isolate, and correct deficiencies.

25X6A

in OCD.

30. Now, in November of 1949, it begins to seem certain that the system has proven its usefulness, and that it may ultimately be made to serve the needs of all intelligence analysts and collectors. There are difficulties, to be sure, and the Classification Scheme comes under sharp criticism every time a new customer commences to use it. Each such new customer is likely to argue that that part of the scheme applying to his own specialty should be entirely revised by himself, and expanded

Next 1 Page(s) In Document Exempt

31. Present status. Growth of the Abstract Files is shown in the diagram on the preceding page, and the sharp acceleration in recent months is cause for concern about ways and means of handling the volume. 1,528 abstracts were put into the files in August, 2,078 in September, and 2,966 in October. Production by the [redacted] seems to be levelling off at about 300 per month, by the Greece desk at about 400, and by OCD itself at about 450. But OSI is still growing, and its production jumped from 478 in August to 1,052 in September and 1,792 in October. We have no guide to assist us in estimating what total volume would be if all desks of the production offices contributed, but it is certain that it would greatly exceed present OCD strength to type and process the abstracts.

25X6A

32. Objectives of the abstracting system may be summarized as follows:

(a) To provide for all authorized readers highly competent summaries or abstracts of the most important intelligence materials in CIA files.

(b) To make the daily output of each specialist, working with materials which are of prime interest to him, available for the use of all other analysts whose interests overlap

25X6A

[redacted]

(c) To provide faster and more economical service for all who call upon the Library for documents; this by giving them lists of titles bearing upon their problems in which the important documents are emphasized and summarized. Thus the user will make faster and better selection of the documents he needs and, there being fewer of them, the Library will make faster delivery.

(d) To determine which documents in CIA files may best be jettisoned to make room for fresh inflow. When the system is in full operation it will be possible to assume that documents selected by analysts for abstracting must be held, while those not so selected may be destroyed after a reasonable period of time.

33. Ways and Means. There is general agreement among all who are using the abstracting system that it has proved its worth, and that it should be extended as rapidly as possible. This will require additional hands to type and code the abstracts, and these additional hands might be placed either in OCD or in the production offices themselves. OCD's five full-time abstract typists are now handling a volume of nearly 3,000 a month. About 30 apiece per working day seems to be about the maximum, as it is a ticklish job to use an electric typewriter on a multigraph sheet only 3 inches wide.

34. Typing by the production offices has been proposed by the Agency's Management Officer, and he has now installed this system in OSI. The advantage hoped for from this expedient is that it will lower costs by avoiding the duplication which now exists when abstracts are first typed by the production office on 8" x 10" transmittal sheets with ordinary typewriters, and then retyped by OCD with electric typewriters on 3-inch multigraph masters. It remains to be seen if a saving will really result from this shift, as it is generally the case that highly specialized typing jobs, using costly machines, are most economically done by a pool of full-time operators. There is also the danger, inherent in decentralization, that the output of the several desks of OSI will lose that uniformity of format and coding which is essential if the abstracts are to be of value to researchers other than those who wrote them. Thus:

Each abstract must carry at its head the bibliographic data - title of report, originating agency, serial number, classification, and number of pages - and it must also carry the Library's coded locator data. The data must be complete, exact, and arranged in a uniform manner. If not, the Library will be unable to find the original documents when they are called for. Experience to date has shown that technical specialists and their clerical helpers tend to be remiss in this essential matter. Careful indoctrination and editing should cure this weakness: if it does not, then many OSI abstracts will have to be retyped before they are reproduced and entered in the files.

Each scientific specialist tends to code in detail for his own specialty but to neglect the more general codes, and the specific geographic codes, which alone make it possible for

his output to be made generally available to other analysts. If this is not corrected then the scientific segments of our abstract files will be of value to the specialists who prepared them, but not to anyone else. A possible solution is to have the code numbers for general reference purposes added by the Library.

These are the major difficulties, and they are not insoluble. A few months should show whether or not daily output per special typewriter can be as high under this system as it can in a central unit.

35. Typing by a central unit probably has advantages which outweigh the minor duplication referred to above. The 8" x 10" transmittal sheets turned out by the production offices can be typed by the offices' regular typists or by the junior professionals who actually write the abstracts, and there is no need for electric typewriters costing over $400 apiece. These sheets themselves are serving a highly useful purpose in ORE, as the carbons are bound together in a chronological file to which the Section Chief turns for the writing of weekly and monthly summaries. They are large enough to leave room for marginal comments of value in training the junior analysts, and for the code numbers which must in any event be transmitted to the Library by one means or another.

Most important, it is possible in a central unit to exercise that degree of supervision which alone can ensure uniformity in format and coding - a uniformity which is absolutely essential if the cards put into the files are to serve the purposes of a central reference system.

HIGHLY SPECIALIZED REFERENCE PROBLEMS

36. There are two types of highly specialized reference problems which cannot be handled by the central reference system, and for which it is necessary to devise solutions adapted to the particular problems involved. Neither type could be solved entirely within the framework of the central reference system without disastrous effects on the latter.

37. The Technical Specialist presents the first of the two types of problems.

Let us assume that he is a medical intelligence officer, and that he is concerned only with medical problems. When he receives a valuable report on public health conditions in Turkey it is necessary that he get it cross-referenced in many different ways. Perhaps it refers to some 40 different diseases, and the treatments employed by [redacted] 25X6A health officers for each of them. In this case he requires that the report be cross-referenced in 40 different ways, in order that it will come rapidly to his hand when next he must write a statement on world-wide incidence of any one of the diseases and the measures employed to combat it. The Library's Classification Scheme is quite inadequate for this purpose, and the report will be of little value to him if it is cross-referenced only under such general headings as "Public Health", "Epidemiology", "Pathology", and the like.

What would happen if the Library attempted to meet the medical officer's needs? It would have to expand the categories in the medical sections of its Classification Scheme into the most minute detail, and it would have to enter 40 different copies of a single abstract in its files. Then, when a [redacted] 25X6A specialist from ORE asked for reports dealing with public health in [redacted] 25X6A he would get a listing on which the same title and abstract appeared 40 different times under as many different sub-headings. Also, if the Library made an all out effort to satisfy the medical specialist, it could scarcely refuse to do the same for other specialists. In no time at all the Intellofax files would be jammed full of duplicate copies of index and abstract faxcards, and the objective of providing a central reference service for all hands would be completely snowed under.

An interim solution has been worked out as follows: The medical officer prepares his abstract, and he applies to it as many different cross-reference code numbers as he desires. These he draws from a specialized Medical Classification Scheme of his own choosing. Each abstract also receives the general cross-reference codes drawn from the Library's Classification Scheme. When OCD receives the abstract it prepares not only the 2 or 3 faxcards needed for the central Abstract Files, but also the 40 extra which the medical officer requires. All cards are punched, those for the medical officer with the extra 6-digit code numbers which he has indicated. The Library cards are then placed in the central files, while the extras are delivered to the medical officer. He handles

these in his own office as he sees fit and, when he finds he needs machine service on them, he can call upon the Machine Division of OCD to sort, select, and arrange his cards by IBM equipment using his own punched code numbers.

Though this is, admittedly, no more than an interim solution, it seems to be working fairly well. The medical officer gets all the cross-reference cards he needs, though he need type each abstract only once. The central reference system benefits from getting its public health documents abstracted by a specialist in the subject, for these abstracts then become available for the use of anyone who needs them. Preparation of the additional faxcards is a negligible charge against manpower, for it is done by machine. The chief flaw in the system is that the medical officer, by having six additional digits punched into his cards, reduces the amount of space available for his abstracts by about one inch. Thus he must compress his abstracts into about 120 instead of 150 words.

Management personnel will doubtless perceive what is to them an additional flaw in the system - i.e. that it will doubtless lead to fairly extensive file-keeping by the medical officer himself. Administrators and researchers are always at loggerheads on this question: the former invariably try to administrate the researcher's files out of existence, while the latter retorts that he knows better how he should perform his job than do the administrators. "If not", he asks, "why don't the management people take over the job of producing medical intelligence?"

OCD is a bystander in this inevitable dispute, but it wishes to develop a central reference system which will satisfy 90% of the requests which it receives. It would be unable to do so if it filled its central files with thousands of minutely categorized duplicate cards on technical subjects, and it is inclined to agree with the research man that he knows best what he needs. If the administrator-folk gain the upper hand in the dispute then perhaps we could satisfy them by setting up separate technical files located in OCD apart from the central files. It is true that this would not save any space or filing equipment, and that it would merely make life a little harder for the technical specialists, but it would make it appear on paper that there was no "duplicate file-keeping."

38. The Mass Data Problem is the second of the two types which cannot satisfactorily be solved within the framework of a central reference

system. Unlike the job needed by the technical specialist it does not usually require extensive cross-referencing of individual reports, but it does demand the cumulation of thousands or millions of individual items of information, identical in nature but differing in substance, in order that they may be treated by machine methods. If these were thrown into a central reference system they would have the same effect as the technical specialist's abundant cross-references - they would cause the central system to bog down in its tracks. What would happen if the millions of cards handled by the Bureau of the Census were thrown into the catalog system of the Library of Congress? There are many such mass data problems. Each requires the development of particular techniques for handling the data in such manner as to make them produce answers to the particular questions asked of them: hence each is best handled by a separate unit.

The Biographic and Industrial Registers are the two units in CIA so far found necessary for problems of this nature. Each must compile many thousands of information items, must bring together in dossiers the items which describe single individuals or industrial plants, and must be able on request to provide indexes, listings, cumulations, and summaries of their data prepared in a variety of different ways.

It is certain that the Registers' cards could not be thrown into the general reference files without wrecking the objectives of the central reference system, and it does not now seem likely that the Intellofax system could profitably be applied to their problems.